Event detection in social media

An analysis of Tweets from the Turkish Gezi Park Uprising

Executive Summary

This thesis takes an explorative approach to determine to which degree events that occur within in the context of a political uprising can be detected from social media posts with computational methods.

My work is based on two pillars: A comparison of systems and approaches that have been developed in previous research of this topic, and the application of a selected hierarchical clustering method to the case of the Turkish Gezi Park Uprising.

The combined learning from these two steps is critically assessed, showing that event detection in the context of a political uprising is feasible, and that such a system can be used for situational awareness. Clustering methods are the most suitable choice for the detection of events by previously unknown categories. Combined applications of clustering and classification are highly suitable for the practical purpose of detecting events by known categories, whereas graph-based methods offer a more versatile approach for specific applications. On the use-cases side, intelligence services and humanitarian agencies are found to have an interest in event detection methods, which holds ethical implications for the deployment of such a system.

Agenda

1. Research Question
2. Methodology
3. Results
4. Findings

Research Question

01

Base assumptions:

- Social media can play an important role for information dissemination during a crisis (Öztürk & Ayvaz, 2018; Ozturkcan et al., 2017), making it a useful source for analysts who are monitoring the situation.
- Filtering relevant information about incidents from masses of posts is a big data analytics problem (eg. Aggarwal, 2011).
- There are ethical implications when working with AI and personal data (eg. Scherer, 2016).
- Solutions for automated detection have already been attempted (eg. Hua et al., 2013; Weng et al., 2011).
- **It is unclear how such systems are actually used, and whether the output is useful for situational awareness (i.e. personal security risk assessment).**
- **Applicability of a given approach to a new case is not clear, especially considering language differences.**

Research Question

To which degree can events which occurred during the Gezi Park Uprising be detected with computational methods applied to Tweets?

Research Question

To which degree can events which occurred during the Gezi Park Uprising be detected with computational methods applied to Tweets?

Event definition of the Topic Detection and Tracking (TDT) initiative:

“Something that happened at a specific time and place with consequences” (Allan, 2002).

Methodology

02

A combination of qualitative and quantitative methods was applied to answer the research question:

- Introduction to core concepts in data mining and social media analytics, such as topic and community detection
- Comparison of 15 different approaches for detecting events from social media
- Scoring table for selection of the best-suitable approach for the Gezi Uprising case
- Application of chosen method to the case, using the CRISP-DM (Wirth & Hipp, 2000) phases as a guiding framework and Python for the implementation
- Discussion of obtained results to answer the research question

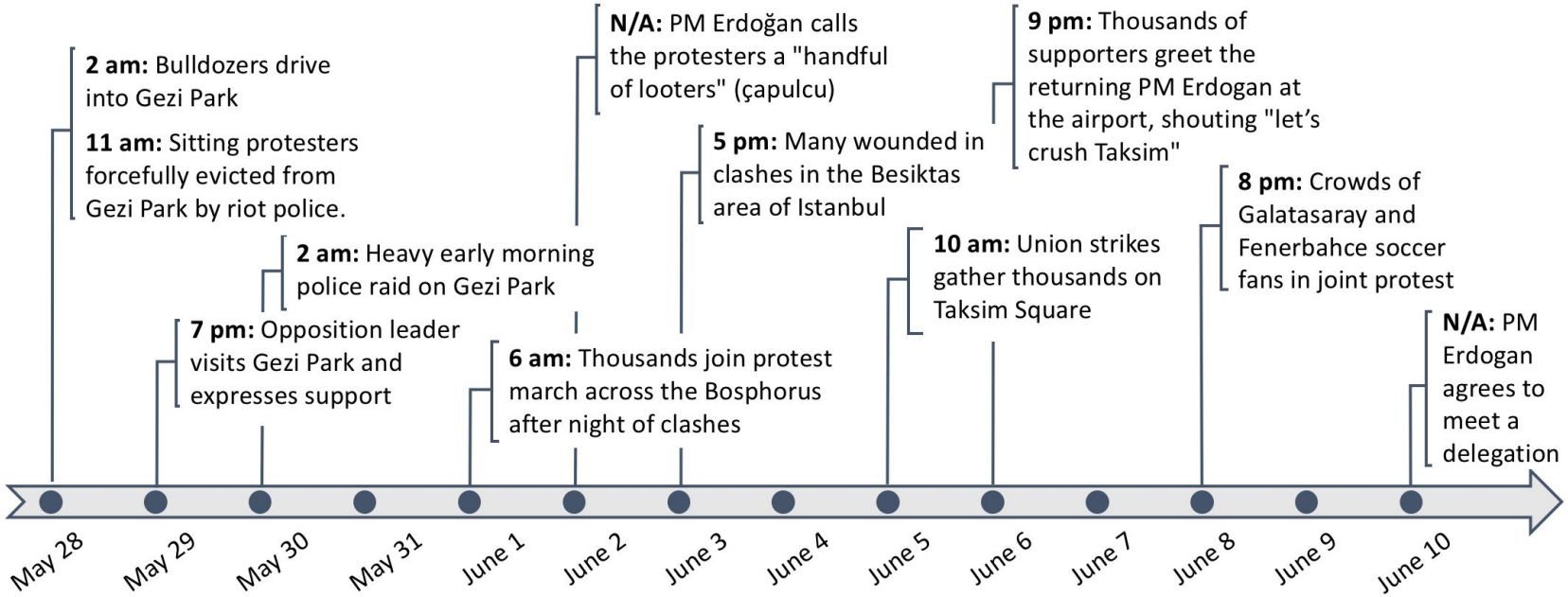
Event-detection methods from previous research were distinguished by **three categories:**

a) Clustering methods
b) Combined applications of clustering and classification methods
c) “Advanced” methods (keywords: graph-based, search techniques, signal processing, computer vision)

Scoring table (excerpt); page 30f

Paper	Method	Mathem. complexity	Transpa-rency	Use of labeled Tweets	Use of geo-location	Suitability assessment for Gezi case
Ranneries et al., 2016	Clustering	Low	High	No	Yes	Low; requires geo-location
Feng et al., 2015	Clustering	Medium	Medium	No	Yes	Low; real-time stream of Tweets used
Yang & Rayz, 2018	Clustering	Low	High	No	No	Low; number of events must be known in advance
Becker, 2011	Clustering	Medium, due to multi-dimensiona l approach	High	No	No	Medium, due to required learning effort and scope

Timeline of the Gezi Park Uprising within the chosen observation period; own illustration:

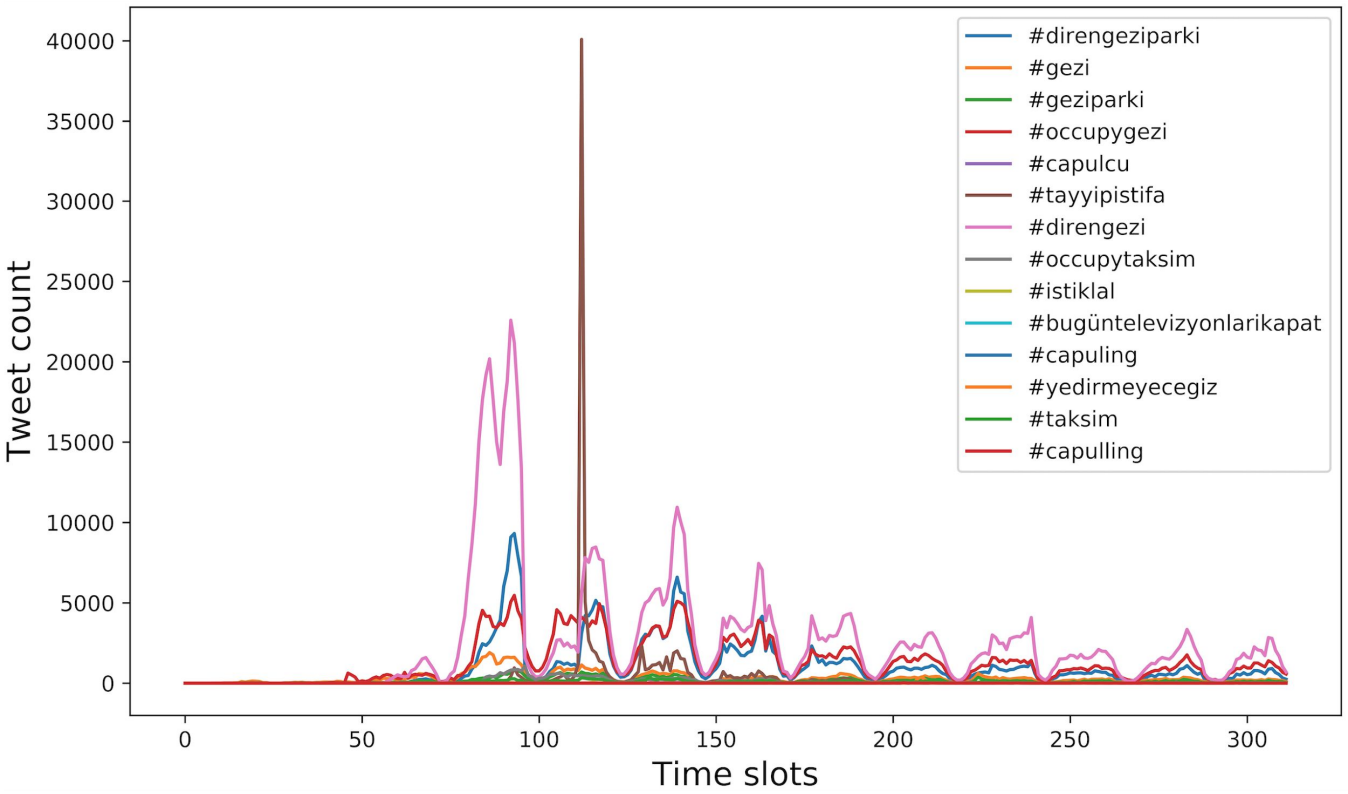


“Lady in red” photo stream depicts the first moment of escalation.

Images taken during the morning police raid on Gezi Park on May 28th (CBS News, Independent)

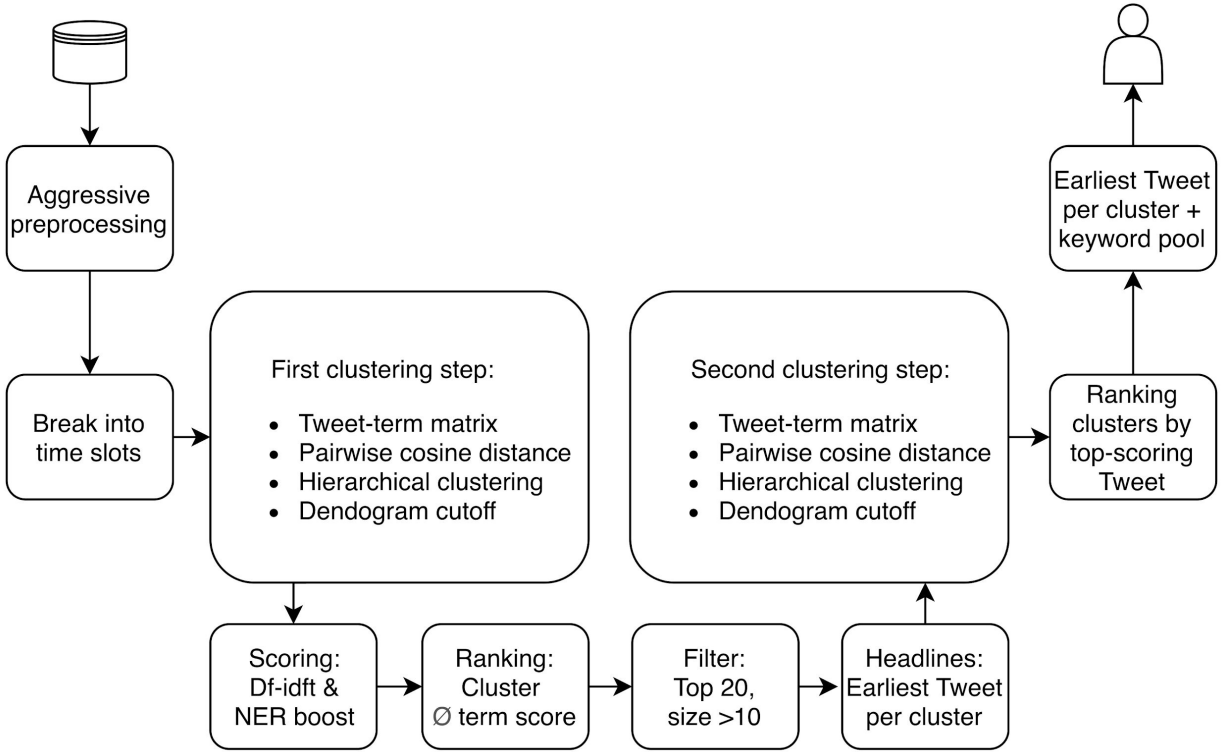
Repeated clashes between protesters and police on Taksim Square, picture from June 11 (Reuters)

Festival-like scenes were reported in Gezi Park, picture from June 1 (Opendemocracy.net)

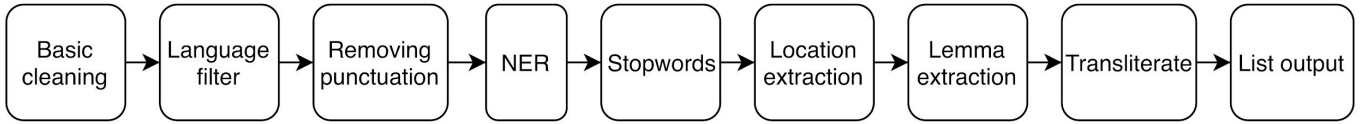


Hourly hashtag usage in 2.2 million Tweets published between May 28 - June 10, 2013 (own illustration)

High-level overview of the Ifrim et al. (2014) method, which was applied to the Gezi Park Uprising data (own illustration):



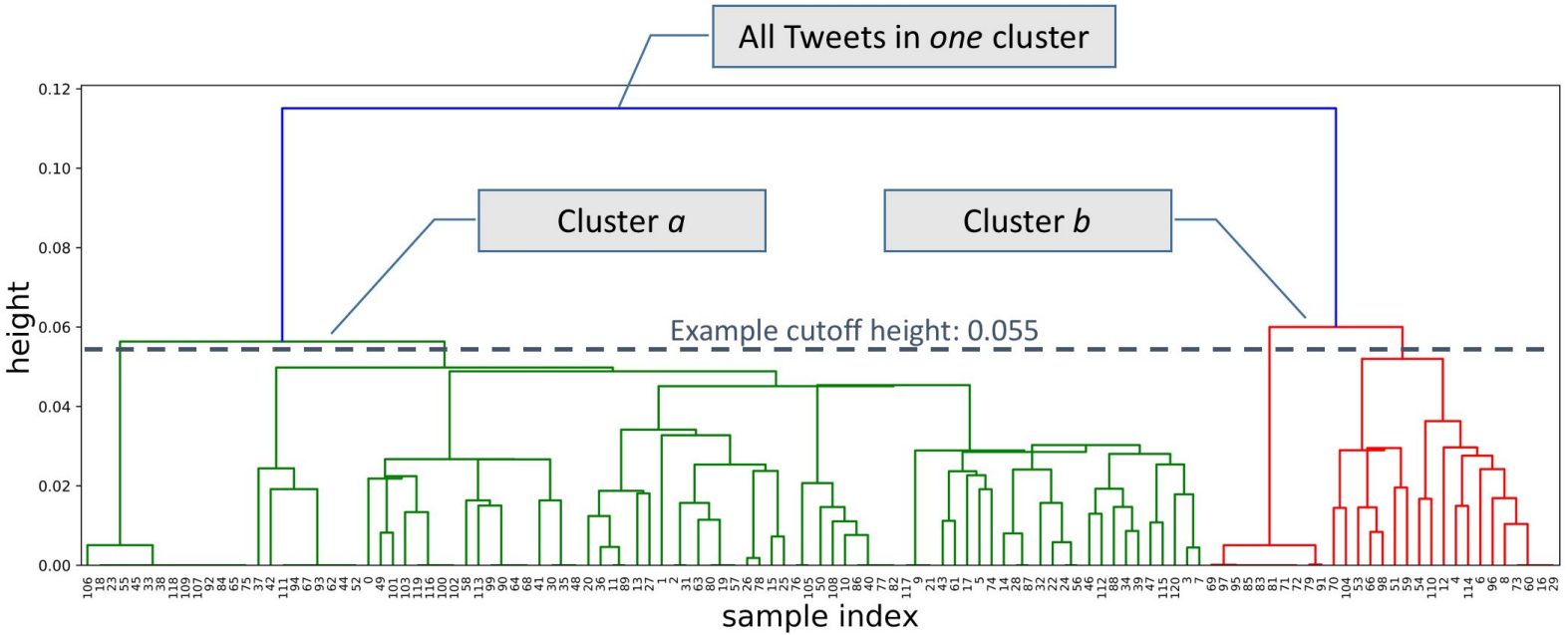
- A highly specific preprocessing pipeline was designed to deal with the Turkish language issues and to optimally prepare the data for modeling:



- The Aiello et al. (2013) "df-idft" method was used to rank clusters based on word occurence in time slots, ensuring that new bursts of keywords trended higher than familiar ones from previous slots.

$$df\text{–}idf_t = \frac{df_i + 1}{\log\left(\frac{\sum_{j=i}^{t} df_{i-j}}{t} + 1\right) + 1}$$

Hierarchical clustering, own concept illustration



Cluster 15 at time 2013-05-30 02:00:07

Tweet text	Gezi Parkı'nı 5 dakika içinde kaybedeceğiz ne yazık ki!
Translation	We will lose the Gezi Park within 5 minutes how sad!
Time	2013-05-30 02:00:07
Reported event: Nightly raid on Gezi Park; heavily armed riot police attempt to drive out the protesters with water cannons, tear gas and batons. (Report by: Hürriyet)	

Tweet text	Ultraslan : Parka polis sokmuyor Gfb : gezi parkini koruyor CARŞI : panzerle polis kovaliyor :)))
Translation	Galatasaray: No police gets into the park Fenerbahce: Protecting the Gezi Park BEŞIKTAŞ: Attacking the police with a tank
Time	2013-06-01 17:03:15
Reported event: Intense clashes between protesters and riot police leave many wounded in the Besiktas district of Istanbul. (Reported by: Reuters)	

Tweet text	3 5 çapulcu olarak Taksim Gezi Parkında eyleme devam ederken, 230'a yakın Harvard mezunu profesör havaalanında başbakanı karşılıyor.
Translation	While 3 to 5 are continuing to showcase their support in Gezi Park as looters, 230 almost-Harvard-graduate-professors are greeting the Prime Minister at the airport.
Time	2013-06-06 22:00:04
Reported event: Thousands of Erdogan's supporters greeting the returning Prime Minister at the airport with shouts "we will die for you" and "let's go crush Taksim". (Reported by: Al Jazeera)	

Results

Performance was evaluated according to **three measures**:

1. Detection rate of pre-selected events of different granularity, magnitude and characteristics.
2. “True positives” rate, which in this case describes the daily rate of system outputs that contained relevant information about reported events.
3. Qualitative assessment of system outputs: comparison of common characteristics of events detected and events not detected.

Number of events clearly detected

22

Pre-selected events which were detected with a clear match and relevant information.

Detection rate (weak + strong matches)

53 %

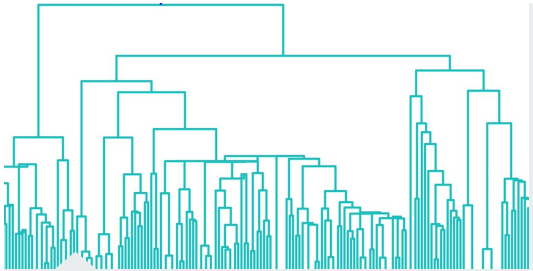
88 % of major events and 29 % of non-major events.

True positives rate

<88 %

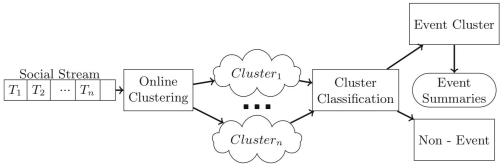
The daily rate of informative outputs ranged between 0 and 88%. The best score of 88 % was achieved both on June 1 and 2.

Findings



Hierarchical clustering is a suitable data mining approach for detecting *major events* in the context of a crisis.

Methods should be selected according to the use case: For practical applications in the crisis context, hybrid approaches offer best results.



Some event-types seem to receive more coverage on social media than others, which has implications for detectability (further research required).

Thanks for your attention!

Q&A

References (1)

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In C. C. Aggarwal (Ed.), Social Network Data Analytics (pp. 1–15). Boston, MA: Springer US.

Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., ... Jaimes, A. (2013). Sensing Trending Topics in Twitter. IEEE Transactions on Multimedia, 15(6), 1268–1282.

Allan, J. (2002). Introduction to Topic Detection and Tracking. In J. Allan (Ed.), Topic Detection and Tracking (Vol. 12, pp. 1–16). Boston, MA: Springer US.

Hua, T., Chen, F., Zhao, L., Lu, C.-T., & Ramakrishnan, N. (2013). STED: Semi-supervised targeted-interest event detection in in twitter. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '13, 1466. Chicago, Illinois, USA: ACM Press.

Ifrim, G., Shi, B., & Brigadir, I. (2014). Event Detection in Twitter using Aggressive Filtering and Hierarchical Tweet Clustering. CEUR Workshop Proceedings, 1150, 33–40.

Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. Telematics and Informatics, 35(1), 136–147.

Ozturkcan, S., Kasap, N., Cevik, M., & Zaman, T. (2017). An analysis of the Gezi Park social movement tweets. Aslib Journal of Information Management, 69(4), 426–440.

References (2)

Scherer, M. U. (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. Harvard Journal of Law & Technology, 29(2), 353.

Weng, J., Yao, Y., Leonardi, E., & Lee, F. (2011). Event Detection in Twitter. Proceedings of the Fifth International Conference on Weblogs and Social Media, 22.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 29–39.

References (images)

Title page image: Retrieved 8/9/2019 from
https://wittymisfitsinc.wordpress.com/2013/06/01/turkish-spring-turkish-citizens-protest-against-fascism-an-islamic-government-and-police-brutality/turkey-protest-photos-occupy-gezi-taksim-gezi-park-45/

Lady in red image stream: Retrieved 8/9/2019 from
https://cbsnews1.cbsistatic.com/hub/i/2013/06/05/ab18f5f9-d25a-11e2-a43e-02911869d855/turkey2.jpg and
https://static.independent.co.uk/s3fs-public/thumbnails/image/2013/06/05/09/Turkey-woman-3.jpg

Page 11: Retrieved 8/9/2019 from https://www.opendemocracy.net/en/turkish-human-rights-and-eu-accession-gezi-park-protests/ and
https://www.reuters.com/article/us-turkey-security-gezi/turkey-escalates-crackdown-on-dissent-six-years-after-gezi-protests-idUSKCN1R00EN

Page 20 (man with flag): Retrieved 8/9/2019 from https://ichef.bbc.co.uk/news/660/media/images/82649000/jpg/_82649906_turkey.jpg

Page 20 (activity diagram): Panagiotou, N., Katakis, I., & Gunopulos, D. (2016). Detecting Events in Online Social Networks: Definitions, Trends and Challenges. In S. Michaelis, N. Piatkowski, & M. Stolpe (Eds.), Solving Large Scale Learning Tasks. Challenges and Algorithms (Vol. 9580, pp. 42–84). Cham: Springer International Publishing.